Theoretical Computer Science 411 (2010) 2632–2646

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Prequential randomness and probability

Vladimir Vovk[a,*], Alexander Shen[b]

[a] Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, England, United Kingdom

[b] Laboratoire d'Informatique Fondamentale, Centre de Mathématiques et Informatique, 39 rue Joliot-Curie, F-13453 Marseille Cedex 13, France

ARTICLE INFO

Keywords:
Algorithmic randomness
Farthingale
Prequential statistics

ABSTRACT

This paper studies Dawid's prequential framework from the point of view of the algorithmic theory of randomness. Our first main result is that two natural notions of randomness coincide. One notion is the prequential version of the measure-theoretic definition due to Martin-Löf, and the other is the prequential version of the game-theoretic definition due to Schnorr and Levin. This is another manifestation of the close relation between the two main paradigms of randomness. The algorithmic theory of randomness can be stripped of its algorithmic aspect and still give meaningful results; the measure-theoretic paradigm then corresponds to Kolmogorov's measure-theoretic probability and the game-theoretic paradigm corresponds to game-theoretic probability. Our second main result is that measure-theoretic probability coincides with game-theoretic probability on all analytic (in particular, Borel) sets.

1. Introduction

We consider the following on-line learning protocol:

PROBABILITY FORECASTING OF BINARY OBSERVATIONS
FOR $n = 1, 2, \dots$:
  Learner announces $p_n \in [0, 1]$.
  Reality announces $y_n \in \{0, 1\}$.
END FOR.

Intuitively, p_n is Learner's subjective probability that $y_n = 1$ after having observed $y_1, \dots, y_{n-1}$ and taking account of all other relevant information available at the time of issuing the forecast. We will refer to p_n as *forecasts* and to y_n as *outcomes*.

When can we say that Learner is doing a good job of forecasting? Or as we shall say, when is the sequence of outcomes $(y_1, y_2, \dots)$ "random" with respect to the sequence of forecasts $(p_1, p_2, \dots)$? (We further abbreviate this by saying that the sequence $(p_1, y_1, p_2, y_2, \dots)$ containing both forecasts and outcomes is random.) This paper demonstrates the equivalence of two superficially quite different answers to this question.

1.1. Standard theory

The simplest, and well studied, situation is where the forecasts are produced as conditional probabilities from a computable probability distribution P on $\{0, 1\}^\infty$: p_n is the conditional probability according to P that $y_n = 1$ given $y_1, \dots, y_{n-1}$. In this case it is natural to talk about the randomness of $(y_1, y_2, \dots)$ with respect to P rather than with respect to $(p_1, p_2, \dots)$.

* Corresponding author. Tel.: +44 1784 443426; fax: +44 1784 439786.
E-mail addresses: vovk@cs.rhul.ac.uk (V. Vovk), alexander.shen@lif.univ-mrs.fr (A. Shen).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2010.04.002

The two fundamental paradigms of randomness are measure-theoretic and game-theoretic. The first paradigm is usually referred to as typicality, and various flavours and modifications of the second paradigm are referred to as unpredictability, stochasticity, chaoticity, and incompressibility (see below). This terminology, however, would be counterintuitive in the general framework of this paper. (To some degree this is also true about the standard term "random", but the reader is perhaps already accustomed to the technical meaning of this term being different from its everyday meaning.)

The game-theoretic notion of randomness is based, as can be guessed, on the idea of gaming: a sequence of outcomes is regarded random if there is no computable way to become infinitely rich betting on its elements. Similarly, the measure-theoretic notion is based on the idea of measure: a sequence is regarded random if there is no computable way to specify a set of measure zero containing this sequence. The standard game-theoretic definition of randomness is due to Schnorr [27] and Levin ([18], Theorem 3), and the standard measure-theoretic definition of randomness is due to Martin-Löf [21]. Schnorr [27] and Levin [18] established the equivalence between the two definitions.

The measure-theoretic definition of randomness with respect to a non-computable probability distribution P was first given by Levin [18] and later developed, modified, and applied in, e.g., [19,11,36,12,26].

Remark 1. The terms "typicality", "chaoticity", and "stochasticity" were used in [17]; "stochasticity" was introduced in Kolmogorov's earlier papers, and "typicality" and "chaoticity" were introduced in [17] itself. Modern literature often talks about "unpredictability" and "incompressibility" (see, e.g., [1], Chapter 1). As we said, typicality is synonymous with what we call measure-theoretic randomness, and the other four terms are various versions of game-theoretic randomness. Unpredictability is synonymous with our game-theoretic randomness. Stochasticity is the game-theoretic notion of randomness based on von Mises's idea of subsequence selection rules, which Ville [31] showed to be inadequate in some important respects. The synonyms "chaoticity" and "incompressibility" require that the algorithmic complexity of initial fragments of the sequence should be close to its trivial upper bound. If the complexity is defined as the minus logarithm of the *a priori* semimeasure, this is the same as unpredictability. However, other notions of complexity (such as plain, prefix, and monotonic) have also been considered, and in this case chaoticity/incompressibility is sometimes regarded as a third, information-theoretic, paradigm, based on coding; but in any case, this third paradigm is very close to the game-theoretic one, as the connections between coding and gambling are straightforward and well understood (see, e.g., [15,4], Chapter 6). Levin's ([18], Theorem 3) representation of game-theoretic randomness is in terms of complexity (the chaoticity/incompressibility approach) and Schnorr's [27] is in terms of martingales (the unpredictability approach).

1.2. Prequential framework for randomness

The standard definitions of randomness mentioned in the previous subsection depend on knowing Learner's probability model P. Our forecasting protocol, however, only involves the realized forecasts p_n, which are not assumed to be derived from any P. This feature of the protocol greatly extends its area of application, allowing forecasts produced "on the fly". Dawid's prequential principle ([5]; it is called "M2" in [6] and "weak prequential principle" in [7,8]) says that our evaluation of the quality of the forecasts $p_1, p_2, \ldots$ in light of the observed outcomes $y_1, y_2, \ldots$ should not depend on Learner's model P even if it exists and is known.

The first definition of randomness fully respecting the prequential principle was proposed by Dawid [6]. Dawid's definition, however, was based on von Mises's idea of subsequence selection rules. Dawid ([6], Section 13.2) also gave a brief description of a prequential definition based on Ville's martingales, but did not elaborate on it. Chernov et al. [2] give the details of the martingale definition in the case where the forecasts are only allowed to take values from a finite set. This paper provides the details of the general martingale definition, which belongs to the game-theoretic paradigm. It also gives a Bayesian definition of measure-theoretic randomness. Its main mathematical result says that the notions of measure-theoretic randomness (called measure-randomness for brevity) and of game-theoretic randomness (called game-randomness) coincide in the prequential framework.

1.3. This paper

In the following two sections we will introduce the two notions of randomness of a sequence of forecasts and observations. Intuitively, $(p_1, y_1, p_2, y_2, \ldots)$ is random if the p_n are good predictions of y_n; slightly more precisely, if there is no computable way to detect inadequacy of p_n. The first definition, in Section 2, belongs to the game-theoretic paradigm and the second, in Section 3, to the measure-theoretic paradigm. The equivalence between the two definitions, which is the first main result of the paper, is stated in Section 4 as Theorem 1 and proved in Section 5. Section 6, which is new as compared to the conference version [35] of this paper, casts Theorem 1 in terms of two kinds of prequential probability, game-theoretic and measure-theoretic. Finally, in Section 7 we prove that the two kinds of prequential probability coincide on all analytic sets. This is our second main result. It was stated in the conference paper [34] (and so this paper is a joint journal version for [35,34]).

1.4. Some notation and definitions

The set of all natural (i.e., positive integer) numbers is denoted by $\mathbb{N}$, $\mathbb{N} := \{1, 2, \ldots\}$; $\overline{\mathbb{N}}_0$ is $\mathbb{N}$ extended by adding ∞ and 0. As always, $\mathbb{Q}$ and $\mathbb{R}$ are the sets of all rational and real numbers, respectively.

Let $\Omega := \{0,1\}^\infty$ be the set of all infinite binary sequences and $\Omega^\diamond := \{0,1\}^*$ be the set of all finite binary sequences. Set $\Pi := ([0,1] \times \{0,1\})^\infty$ and $\Pi^\diamond := ([0,1] \times \{0,1\})^*$. The empty element (sequence of length zero) of both $\Omega^\diamond$ and $\Pi^\diamond$ will be denoted by Λ. In our applications, the elements of Ω and $\Omega^\diamond$ will be sequences of outcomes (infinite or finite), and the elements of Π and $\Pi^\diamond$ will be sequences of forecasts and outcomes (infinite or finite). The set Π will sometimes be referred to as the *prequential space*.

For $x \in \Omega^\diamond$, let $\Gamma_x \subseteq \Omega$ be the set of all infinite continuations of x. Similarly, for $x \in \Pi^\diamond$, $\Gamma_x \subseteq \Pi$ is the set of all infinite continuations of x. For each $\omega = (y_1, y_2, \ldots) \in \Omega$ and $n \in \mathbb{N}$, set $\omega^n := (y_1, \ldots, y_n)$. Similarly, for each $\pi = (p_1, y_1, p_2, y_2, \ldots) \in \Pi$ and $n \in \mathbb{N}$, set $\pi^n := (p_1, y_1, \ldots, p_n, y_n)$.

In Appendix and some proofs we will be using the following notation, for $n \in \mathbb{N}$: $\Omega^n := \{0,1\}^n$ is the set of all finite binary sequences of length n; $\Omega^{\le n}$ (resp. $\Omega^{\ge n}$) is the set of all finite binary sequences of length at most (resp. at least) n; $\Pi^n := ([0,1] \times \{0,1\})^n$; $\Pi^{\ge n} := \bigcup_{i=n}^\infty ([0,1] \times \{0,1\})^i$.

For understanding the intuitive meaning of our statements, the following intuitive idea of lower semicomputability will suffice: a function $f : X \to \mathbb{R} \cup \{\infty\}$ is lower semicomputable if there is an algorithm that, for all $x \in X$ and $r \in \mathbb{R}$, will eventually tell us that $f(x) > r$ if this inequality is indeed true. (Lower semicomputable functions are not necessarily computable as the algorithm can work arbitrarily long.) Understanding the proofs requires precise definitions, as given in, e.g., Appendix.

2. Game-randomness

A *farthingale* is a function $V : \Pi^\diamond \to [-\infty, \infty]$ satisfying

$$V(p_1, y_1, \ldots, p_{n-1}, y_{n-1}) = (1 - p_n) V(p_1, y_1, \ldots, p_{n-1}, y_{n-1}, p_n, 0) + p_n V(p_1, y_1, \ldots, p_{n-1}, y_{n-1}, p_n, 1) \tag{1}$$

for all n and all $(p_1, y_1, p_2, y_2, \ldots) \in \Pi$; the products 0∞ and $0(-\infty)$ are defined to be 0. If we replace "=" by "≥" in (1), we get the definition of *superfarthingales*. These are prequential versions of the standard notions of martingale and supermartingale, and in our terminology we follow [8]. We will be interested mainly in non-negative farthingales and superfarthingales.

The value of a farthingale can be interpreted as the capital of a gambler betting according to the odds announced by Learner. In the case of superfarthingales, the gambler is allowed to throw away part of his capital.

Lemma 1. *Let $\mathcal{V}$ be the class of all non-negative lower semicomputable superfarthingales V with initial value $V(\Lambda) = 1$. There exists a largest superfarthingale in $\mathcal{V}$ to within a constant factor. In other words, there exists a superfarthingale $V \in \mathcal{V}$ such that, for any other superfarthingale $V' \in \mathcal{V}$, there exists a constant $C > 0$ such that, for any $x \in \Pi^\diamond$, $V(x) \ge V'(x)/C$.*

Proof. Fix a universal computable sequence of lower semicomputable functions $f_1 : \Pi^\diamond \to [0,\infty], f_2 : \Pi^\diamond \to [0,\infty], \ldots$ (see Lemma 13 in Appendix). It is easy to construct a new computable sequence of lower semicomputable functions $f'_1, f'_2, \ldots$ such that each of f'_l is a superfarthingale in $\mathcal{V}$ and that $f'_l = f_l$ whenever f_l is already in $\mathcal{V}$, $l \in \mathbb{N}$. Then $\sum_{l=1}^\infty 2^{-l} f'_l$ will be a largest, to within a constant factor, superfarthingale in $\mathcal{V}$. □

Let us fix a largest, to within a constant factor, superfarthingale U in $\mathcal{V}$ and call it the *universal* superfarthingale.

Definition 1. A sequence $\pi \in \Pi$ is called *game-random* if $U(\pi^n)$ stays bounded as $n \to \infty$.

The following lemma gives an equivalent definition of game-random sequences.

Lemma 2. *A sequence $\pi \in \Pi$ is game-random if and only if $U(\pi^n)$ does not tend to infinity as $n \to \infty$.*

Proof. Following the proof of Lemma 3.1 in [29], we can construct a superfarthingale $V \in \mathcal{V}$ such that $\liminf_{n\to\infty} V(\pi^n) = \infty$ whenever $\sup_n U(\pi^n) = \infty$. (Therefore, $\liminf_{n\to\infty} U(\pi^n) = \infty$ whenever $\sup_n U(\pi^n) = \infty$.) Indeed, for each $m \in \mathbb{N}$, the function $U^m : \Pi^\diamond \to [0,\infty)$ defined by

$$U^m(x) := \begin{cases} 2^m & \text{if } U(y) > 2^m \text{ for some prefix } y \text{ of } x \\ U(x) & \text{otherwise} \end{cases}$$

is a superfarthingale; it is clear that it is lower semicomputable and so belongs to $\mathcal{V}$. Since $U^1, U^2, \ldots$ is a computable sequence of lower semicomputable functions, we can set

$$V := \sum_{m=1}^\infty 2^{-m} U^m. \quad \square$$

3. Measure-randomness

We can also adapt the standard measure-theoretic definition of randomness to the prequential framework. First we give an informal version of the definition.

A *forecasting system* is a function $\phi : \Omega^\diamond \to [0,1]$. Let Φ be the set of all forecasting systems. For each $\phi \in \Phi$ there exists a unique probability measure $\mathbb{P}_\phi$ on Ω such that, for each $x \in \Omega^\diamond$, $\mathbb{P}_\phi(\Gamma_{x1}) = \phi(x)\mathbb{P}_\phi(\Gamma_x)$. (In other words, such that $\phi(x)$ is a version of the conditional probability, according to $\mathbb{P}_\phi$, that x will be followed by 1.) The notion of a forecasting system is close to that of a probability measure on Ω: the correspondence $\phi \mapsto \mathbb{P}_\phi$ becomes an isomorphism if we only consider forecasting systems taking values in the open interval $(0,1)$ and probability measures taking positive values on the sets Γ_x, $x \in \Omega^\diamond$.

Informally, we say that a sequence $\omega \in \Omega$ is *measure-random* with respect to a forecasting system ϕ if it is random in the sense of Martin-Löf [21] with respect to $\mathbb{P}_\phi$ when ϕ is given as an oracle. We will formalize "given as an oracle" using some simplest notions of effective topology (see Appendix). The following definition is a version of Levin's "uniform test of randomness" [18,19,12].

Definition 2. A *uniform test of randomness* is a lower semicomputable function $T : \Omega \times \Phi \to \overline{\mathbb{N}}_0$ such that, for all $\phi \in \Phi$ and all $m \in \mathbb{N}$,

$$\mathbb{P}_\phi\{\omega \in \Omega \mid T(\omega, \phi) \geq m\} \leq 2^{-m}. \tag{2}$$

Intuitively, $T(\omega, \phi)$ is the amount of irregularities (measured in bits, according to (2)) discovered in ω with respect to ϕ. The requirement of lower semicomputability means that the irregularities have to be genuine: a discovery of irregularity can never be undone. We will usually drop the adjective "uniform".

Lemma 3. *There exists a largest, to within an additive constant, test of randomness. In other words, there exists a test of randomness T such that, for any other test of randomness T', there exists a constant C such that, for any $(\omega, \phi) \in \Omega \times \Phi$,*

$$T(\omega, \phi) \geq T'(\omega, \phi) - C.$$

Proof. The proof is similar to the standard one given by Martin-Löf [21]; it will, however, crucially depend on the compactness of Φ, as in [18,12]. For each set $G \subseteq \Omega \times \Phi$ and each $\phi \in \Phi$ we will use the notation

$$G[\phi] := \{\omega \in \Omega \mid (\omega, \phi) \in G\}$$

for the *ϕ-cut* of G. A convenient alternative representation of a test of randomness T is as a computable sequence of nested open sets $G_1 \supseteq G_2 \supseteq \cdots$ in $\Omega \times \Phi$ such that

$$\mathbb{P}_\phi(G_m[\phi]) \leq 2^{-m} \tag{3}$$

for all $\phi \in \Phi$ and $m \in \mathbb{N}$. This alternative representation will be referred to as the *set representation*, as opposed to the original *functional representation*. It is easy to see that the two representations are indeed equivalent: when given T we can set $G_m := \{(\omega, \phi) \mid T(\omega, \phi) \geq m\}$ (the sequence $G_1, G_2, \ldots$ of open sets is then computable by Lemma 14, which is uniform in C), and when given $G_1, G_2, \ldots$ we can set $T(\omega, \phi) := \max\{m \mid (\omega, \phi) \in G_m\}$. Such sequences $G_1, G_2, \ldots$ will also be referred to as *tests of randomness*.

Let $G_{l,m}$ be a universal computable family of sequences of open sets (cf. Lemma 12 in Appendix). Put $G'_{l,m} := \bigcap_{i=1}^{m} G_{l,i}$, so that $G'_{l,m}$ is a computable family of nested sequences of open sets containing all nested computable sequences of open sets. We can further "trim" each $G'_{l,m}$ to $G''_{l,m}$ so that:

- $\mathbb{P}_\phi(G''_{l,m}[\phi]) \leq 2^{-m}$ for all $\phi \in \Phi$;
- $G''_{l,m} = G'_{l,m}$ whenever $\mathbb{P}_\phi(G'_{l,m}[\phi]) < 2^{-m}$ for all $\phi \in \Phi$.

Indeed, let $G'_{l,m} = \cup\{U_k \mid (l, m, k) \in A\}$ be the representation of $G'_{l,m}$ as the union of basic sets, with A recursively enumerable. Fix temporarily l and m. Set $H_K := \cup\{U_k \mid (l, m, k) \in A, k \leq K\}$, so that $H_1, H_2, \ldots$ is a non-decreasing sequence of simple sets whose union is $G'_{l,m}$. Remember that, by (27), $\overline{H_K} \subseteq G'_{l,m}$. We may "quarantine" new H_K until they are "cleared", i.e.,

$$\forall \phi \in \Phi : \mathbb{P}_\phi(\overline{H_K}[\phi]) < 2^{-m} \tag{4}$$

is established. The open set $G''_{l,m}$ is defined as the union of the H_K that are cleared.

Let us check that condition (4) can indeed be eventually established by a computable procedure when it is satisfied. Suppose (4) is satisfied. The set

$$S := \left\{\phi \in \Phi \mid \mathbb{P}_\phi(\overline{H_K}[\phi]) < 2^{-m}\right\}$$

is effectively open, so that we can effectively generate a sequence of basic sets $U'_k \subseteq \Phi$ whose union is S. By the compactness of Φ, already a finite number of U'_k will cover S when $S = \Phi$, and so (4) can be established in a computable manner.

Therefore, we can list all tests of randomness, in the following sense: there is a computable sequence $(G''_{l,m})_{m=1}^{\infty}$, $l = 1, 2, \ldots$, of tests of randomness that contains all "strict" tests of randomness (i.e., those satisfying the required inequality with "$<$" instead of "$\leq$"; any test of randomness G_m can be made strict by redefining $G_m := G_{m+1}$, $m = 1, 2, \ldots$). To obtain a largest test of randomness G_m, it suffices to set

$$G_m := \bigcup_{l=1}^{\infty} G''_{l,m+l}.$$

Indeed, the computability of the sequence of open sets G_m is obvious,

$$\mathbb{P}_\phi(G_m[\phi]) \leq \sum_{l=1}^{\infty} \mathbb{P}_\phi(G''_{l,m+l}[\phi]) \leq \sum_{l=1}^{\infty} 2^{-m-l} = 2^{-m}, \quad \forall \phi \in \Phi, \forall m \in \mathbb{N},$$

and, for each $l \in \mathbb{N}$,

$$T(\omega, \phi) = \max\{m \mid (\omega, \phi) \in G_m\} \geq \max\{m \mid (\omega, \phi) \in G''_{l,m+l}\} = T_l(\omega, \phi) - l, \quad \forall (\omega, \phi) \in \Omega \times \Phi,$$

where T is the functional representation of the test $(G_m)_{m=1}^{\infty}$ and T_l is the functional representation of the test $(G''_{l,m})_{m=1}^{\infty}$. □

Let us fix a largest, to within an additive constant, test of randomness T and call it the *universal test of randomness*. A sequence $\omega \in \Omega$ is said to be *measure-random with respect to* $\phi \in \Phi$ if $T(\omega, \phi) < \infty$.

Definition 3. We say that $\pi = (p_1, y_1, p_2, y_2, \ldots) \in \Pi$ is *measure-random* if there exists a forecasting system ϕ such that $(y_1, y_2, \ldots)$ is measure-random with respect to ϕ and ϕ *agrees* with π, in the sense that $p_n = \phi(y_1, \ldots, y_{n-1})$ for all $n \in \mathbb{N}$.

4. Equivalence of the two notions of randomness

Theorem 1. *A sequence $\pi \in \Pi$ is game-random if and only if it is measure-random.*

This theorem will be proved in the next section. The proof will be based on Levin's [18] ideas (see also [12]). A related result is Theorem 7 in [2], which is technically much simpler but uses a less natural definition.

The philosophical significance of Theorem 1 is that it establishes the equivalence of the purely prequential and Bayesian viewpoints in the framework of the algorithmic theory of randomness. The definition of measure-randomness is Bayesian, in that Learner is modelled as a coherent decision maker, computing his forecasts by conditioning a probability measure; rejecting the forecasts is the same as rejecting all probability measures that could have produced those forecasts. The definition of game-randomness is purely prequential, in that it does not postulate any probability measures behind the forecasts; the latter are used for testing directly.

A simple corollary of Theorem 1 is the following observation:

Corollary 1. *Let ϕ be a computable forecasting system such that $\phi(x) \in (0, 1)$ for all $x \in \Omega^{\diamond}$. A binary sequence $(y_1, y_2, \ldots)$ is random with respect to $\mathbb{P}_{\phi}$ in the sense of Martin-Löf if and only if the sequence $(p_1, y_1, p_2, y_2, \ldots)$ is game-random (equivalently, measure-random), where $p_n := \phi(y_1, \ldots, y_{n-1})$, $n \in \mathbb{N}$.*

Therefore, the prequential notions of game-randomness and measure-randomness generalize Martin-Löf's notion of randomness. Corollary 1 generalizes Theorem 10 in [2].

Remark 2. Notice that we have never assumed that the past observations $y_1, \ldots, y_{n-1}$ are the only information available to Learner when choosing the forecast p_n for the next outcome y_n. Learner is allowed to (and typically does) use all kinds of "side information" in addition to the past observations. It is easy to extend all our definitions and results to the case where some of this side information, x_n, is also known to the gambler. (As in [6], Section 9, and [2].) As an example, the definition of a farthingale, (1), becomes

$$\begin{aligned} V(x_1, p_1, y_1, \ldots, x_{n-1}, p_{n-1}, y_{n-1}) \;&=\; (1 - p_n)V(x_1, p_1, y_1, \ldots, x_{n-1}, p_{n-1}, y_{n-1}, x_n, p_n, 0) \\ &\quad + p_n V(x_1, p_1, y_1, \ldots, x_{n-1}, p_{n-1}, y_{n-1}, x_n, p_n, 1). \end{aligned}$$

Remark 3. Since we do not record side information in the main part of this paper, the forecasting systems that we consider are never assumed computable: even if Learner computes each forecast from the past outcomes and the side information, typically the forecast cannot be computed from the past outcomes alone. It is not even obvious that the notion of a forecasting system ϕ as we defined it (a function of past outcomes) is meaningful outside purely automatic forecasting. It involves the following scenario of repeated "local surgeries", along the lines of Pearl [25] (Section 6.2). To elicit the value of the function ϕ on a binary sequence $y_1, \ldots, y_n$, we act as follows. First we wait until Reality produces the first piece of side information x_1 and, in response, Learner produces p_1. Then we perform a "local surgery" replacing Reality's outcome by y_1 (if it is different from y_1). Now Reality produces x_2 and Learner produces p_2. Another local surgery replaces the outcome by y_2, etc. Finally, Learner produces p_n, which is taken to be the value of ϕ on $y_1, \ldots, y_n$. Pearl's approach is sometimes regarded as philosophically questionable (see, e.g., Section 4 of Pearl's response in [24]). However, Theorem 1 shows that it leads to the same notion of prequential randomness (and Theorem 2 will show that it leads to the same notion of prequential probability) as the philosophically immaculate approach of Section 2. This can serve as the justification of Pearl's approach in the prequential framework.

Remark 4. It is easy to see that Theorem 1 fails if in the definition of measure-typicalness we require that ϕ should range over computable forecasting systems. Indeed, take any non-computable sequence $(y_1, y_2, \ldots) \in \Omega$ and consider $\pi := (y_1, y_1, y_2, y_2, \ldots)$ as an element of Π. It is clear that π is game-random (no farthingale can grow on it) but no computable forecasting system agrees with π.

5. Proof of Theorem 1 and Corollary 1

The proof of the theorem will depend on a fundamental result called Ville's inequality. Let ϕ be a forecasting system. A *martingale* with respect to ϕ is a function $V : \Omega^{\diamond} \to [-\infty, \infty]$ satisfying

$$V(x) = (1 - \phi(x))V(x, 0) + \phi(x)V(x, 1) \tag{5}$$

for all $x \in \Omega^\diamond$ (with the same convention $0(\pm\infty) := 0$). If we replace "=" by "≥" (respectively, by "≤") in (5), we get the definition of a *supermartingale* (respectively, *submartingale*) with respect to ϕ.

Proposition 1 (*Ville's Inequality, [31], p. 100*). *If ϕ is a forecasting system, V is a non-negative supermartingale with respect to ϕ with initial value $V(\Lambda) = 1$, and $C > 0$,*

$$\mathbb{P}_\phi \left\{ \omega \in \Omega \mid \sup_n V(\omega^n) \geq C \right\} \leq \frac{1}{C}.$$

Fix $\pi \in \Pi$.

Part "if" of Theorem 1

Suppose π is not game-random. Then $\pi \in G_m$ for all $m \in \mathbb{N}$, where

$$G_m := \left\{ \pi \in \Pi \mid \sup_n U(\pi^n) > 2^m \right\}$$

and U is the universal superfarthingale. For $\phi \in \Phi$ and $\omega = (y_1, y_2, \ldots) \in \Omega$ we set

$$\omega^\phi := (\phi(\Lambda), y_1, \phi(y_1), y_2, \phi(y_1, y_2), y_3, \ldots) \in \Pi;$$

similarly, for $\phi \in \Phi$ and $x = (y_1, \ldots, y_n) \in \Omega^\diamond$ we set

$$x^\phi := (\phi(\Lambda), y_1, \phi(y_1), y_2, \ldots, \phi(y_1, \ldots, y_{n-1}), y_n) \in \Pi^\diamond.$$

The mapping $(\omega, \phi) \mapsto \omega^\phi$ from $\Omega \times \Phi$ to Π is continuous. Therefore, the set

$$G'_m := \left\{ (\omega, \phi) \mid \omega^\phi \in G_m \right\}$$

is open.

Let us check that G'_m is a test of randomness. The computability requirement follows from Lemma 14 (suitably modified; in particular, made uniform in C) in Appendix. Fix $m \in \mathbb{N}$ and $\phi \in \Phi$. To check (3), i.e., $\mathbb{P}_\phi(G'_m[\phi]) \leq 2^{-m}$ in the current notation, notice that the function $U^\phi : \Omega^\diamond \to [0, \infty]$ defined by

$$U^\phi(x) := U\left(x^\phi\right), \quad x \in \Omega^\diamond, \tag{6}$$

is a non-negative supermartingale with respect to ϕ. Now Ville's inequality implies

$$\begin{aligned} \mathbb{P}_\phi(G'_m[\phi]) &= \mathbb{P}_\phi\left\{\omega \in \Omega \mid (\omega, \phi) \in G'_m\right\} = \mathbb{P}_\phi\left\{\omega \in \Omega \mid \omega^\phi \in G_m\right\} \\ &= \mathbb{P}_\phi\left\{\omega \in \Omega \mid \sup_n U^\phi(\omega^n) > 2^m\right\} \leq 2^{-m}, \quad \forall \phi \in \Phi. \end{aligned}$$

Suppose π, assumed to be not game-random, is measure-random. Then there exists $\phi \in \Phi$ such that $\pi = \omega^\phi$ for some ω measure-random with respect to ϕ. Since $\pi \in G_m$, we have $(\omega, \phi) \in G'_m$; since this is true for each $m \in \mathbb{N}$, ω is not measure-random with respect to ϕ, and so we have arrived at a contradiction.

Part "only if" of Theorem 1

Let $G_m = \cup\{U_k \mid (m, k) \in A\}$ be a representation of the universal test of randomness via basic sets, with $A \subseteq \mathbb{N}^2$ a recursively enumerable set. Without loss of generality we can assume that each basic set U_k in this representation has the form $\Gamma_c \times \left\{\phi \in \Phi \mid a(x) < \phi(x) < b(x), \forall x \in \Omega^{\leq n}\right\}$ for some $c \in \Omega^n$, $a, b : \Omega^{\leq n} \to \mathbb{Q}$, and $n \in \mathbb{N}$. Define G'_m to be the set of all $(p_1, y_1, p_2, y_2, \ldots) \in \Pi$ such that $((y_1, y_2, \ldots), \phi) \in G_m$ for all ϕ that agree with $(p_1, y_1, p_2, y_2, \ldots)$.

The compactness of Φ easily implies that each set $G'_m \subseteq \Pi$ is open. Indeed, suppose $\pi = (p_1, y_1, p_2, y_2, \ldots) \in G'_m$. For each $\phi \in \Phi$, either ϕ disagrees with π or $((y_1, y_2, \ldots), \phi) \in G_m$. In both cases there is a neighbourhood O'_ϕ of π and a neighbourhood O''_ϕ of ϕ such that either all elements of O''_ϕ disagree with all elements of O'_ϕ or $((y'_1, y'_2, \ldots), \phi') \in G_m$ for all $(p'_1, y'_1, p'_2, y'_2, \ldots) \in O'_\phi$ and all $\phi' \in O''_\phi$. Since Φ is compact, there is a finite set $\phi_1, \ldots, \phi_J$ such that $\bigcup_{j=1}^J O''_{\phi_j} = \Phi$. We can see that the neighbourhood $\bigcap_{j=1}^J O'_{\phi_j}$ of π is a subset of G'_m.

Essentially the same argument shows that the G'_m form a computable sequence of open sets. Let us show that there exists a non-negative superfarthingale V_m with initial value 2^{-m} or less that eventually exceeds 1 on each sequence in G'_m. (In this sense G'_m form a prequential test of randomness.)

Let $G'_m = \cup\{U_k \mid (m, k) \in A\}$ be a representation of G'_m via basic sets, where $A \subseteq \mathbb{N}^2$ is a recursively enumerable set. Let $A = \bigcup_{i=1}^\infty A_i$ be a representation of A as the union of a computable nested sequence $\emptyset \subset A_1 \subseteq A_2 \subseteq \cdots$ of finite sets. Fix an m. For each $i \in \mathbb{N}$, define a superfarthingale W_i as follows. Let N be so large that, for all $x \in \Pi^N$ and $(m, k) \in A_i$, either

$\Gamma_x \subseteq U_k$ or $\Gamma_x \cap U_k = \emptyset$. (For example, we can set N to the largest n_k in (26) over k such that $(m, k) \in A_i$.) For $n \geq N$ and $x \in \Pi^n$, set

$$W_i(x) := \begin{cases} 1 & \text{if } \Gamma_x \subseteq U_k \text{ for some } k \text{ with } (m,k) \in A_i \\ 0 & \text{otherwise.} \end{cases}$$

After that proceed by backward induction. If $W_i(x)$ is already defined for $x \in \Pi^n$, $n = N, N-1, \ldots, 1$, set, for each $x \in \Pi^{n-1}$,

$$W_i(x) := \sup_{p \in [0,1]} \big((1-p)W_i(x, p, 0) + pW_i(x, p, 1)\big). \tag{7}$$

It is clear that W_i is a superfarthingale that does not depend on the choice of N.

We will need to establish several properties of W_i. First, it is lower semicontinuous. Indeed, there is an N (e.g., the largest n_k in (26) over $(m, k) \in A_i$) such that $W_i(x)$ is lower semicontinuous when x is restricted to Π^n with $n \geq N$. (It will be even lower semicomputable when x is restricted to $\Pi^{\geq N}$.) And the operation sup preserves lower semicontinuity:

Lemma 4. *If a function $f : X \times Y \to \mathbb{R}$ defined on the product of topological spaces X and Y is lower semicontinuous, then the function $x \in X \mapsto g(x) := \sup_{y \in Y} f(x, y)$ is also lower semicontinuous.*

Proof. It suffices to notice that, for each $c \in \mathbb{R}$, $\{x \mid g(x) > c\} = \{x \mid \exists y : f(x, y) > c\}$, and projections of open sets are open. $\square$

The lower semicontinuity of W_i implies its lower semicomputability: indeed, we can restrict p to $\mathbb{Q} \cap [0, 1]$ in (7).

Let us check that $W_i(\Lambda) \leq 2^{-m}$. Suppose that, on the contrary, $W_i(\Lambda) > 2^{-m}$. Construct a forecasting system ϕ as follows. (The words such as "construct" and "choose" are not intended to imply computability: there are no computability restrictions in this paragraph.) For each $x \in \Omega^n$, $n = 0, 1, \ldots, N-1$, choose $\phi(x)$ such that

$$\begin{aligned}(1-\phi(x))W_i(x^\phi, \phi(x), 0) + \phi(x)W_i(x^\phi, \phi(x), 1) \\ \geq \sup_{p \in [0,1]} \big((1-p)W_i(x^\phi, p, 0) + pW_i(x^\phi, p, 1)\big) - \epsilon/N = W_i(x^\phi) - \epsilon/N,\end{aligned}$$

where $\epsilon > 0$ satisfies $W_i(\Lambda) > 2^{-m} + \epsilon$. For each $x \in \Omega^{\geq N}$, define $\phi(x)$ arbitrarily, say $\phi(x) := 0$. Since $\omega^\phi \notin G'_m$ for all $\omega \notin G_m[\phi]$, we have $W_i^\phi(\omega^N) = 0$ for all $\omega \notin G_m[\phi]$. Combining the fact that

$$\left\{\omega \mid W_i^\phi(\omega^N) = 1\right\} \subseteq G_m[\phi]$$

with the fact that the function $x \in \Omega^\diamond \mapsto S(x) := W_i^\phi(x) + \epsilon n/N$, where n is the length of x, is a submartingale with respect to ϕ, we obtain

$$\begin{aligned}\mathbb{P}_\phi(G_m[\phi]) &\geq \mathbb{P}_\phi\left\{\omega \mid W_i^\phi(\omega^N) = 1\right\} = \mathbb{E}_\phi W_i^\phi(\omega^N) = \mathbb{E}_\phi(S(\omega^N) - \epsilon) \\ &\geq S(\Lambda) - \epsilon = W_i^\phi(\Lambda) - \epsilon = W_i(\Lambda) - \epsilon > 2^{-m},\end{aligned} \tag{8}$$

where $\mathbb{E}_\phi$ stands for the expectation of a function of $\omega \in \Omega$ with respect to $\mathbb{P}_\phi$. The inequality between the extreme terms of (8) fails by the definition of a test of randomness.

Define $V_m(x) := \sup_i W_i(x)$, $x \in \Pi^\diamond$, to be the limit of the non-decreasing sequence of superfarthingales W_i. It is clear that V_m is also a superfarthingale (even satisfying

$$V_m(x) = \sup_{p \in [0,1]} \big((1-p)V_m(x, p, 0) + pV_m(x, p, 1)\big),$$

according to (7)) and $V_m(\Lambda) \leq 2^{-m}$. Set $V := \sum_{m=1}^\infty V_m$; this is a lower semicomputable superfarthingale with initial value $V(\Lambda) \leq 1$ (so that $V \in \mathcal{V}$ if we redefine $V(\Lambda) := 1$).

Now it is easy to finish the proof of the theorem. Suppose that π is not measure-random. Then $\pi \in G'_m$ for all $m \in \mathbb{N}$. Then $V(\pi^n) \to \infty$ as $n \to \infty$, and so π is not game-random.

Corollary 1

Fix a sequence $\omega = (y_1, y_2, \ldots) \in \Omega$, and set $\pi := \omega^\phi \in \Pi$. We will prove the equivalence of the game-randomness of π and Schnorr and Levin's reformulation of Martin-Löf randomness of ω with respect to $\mathbb{P}_\phi$. Remember that Schnorr and Levin's reformulation is that the universal lower semicomputable supermartingale with respect to ϕ is bounded on ω^n, $n \to \infty$; we will refer to this property as the *Schnorr–Levin randomness* of ω with respect to ϕ.

Suppose π is not game-random. Then $U(\pi^n) \to \infty$ as $n \to \infty$, where U is the universal superfarthingale. Then $U^\phi(\omega^n) \to \infty$, and since U^ϕ is a lower semicomputable supermartingale with respect to ϕ, ω is not Schnorr–Levin random.

Now suppose ω is not Schnorr–Levin random with respect to ϕ. Let S be the universal lower semicomputable supermartingale with respect to ϕ. We can transform S into a superfarthingale V by multiplying it by the likelihood ratio:

$$V(p_1, y_1, \ldots, p_n, y_n) := S(y_1, \ldots, y_n) \frac{\mathbb{P}_\phi(\Gamma_{(y_1,\ldots,y_n)})}{\prod_{i=1}^{n} |p_i + y_i - 1|}$$

(the expression $|p_i + y_i - 1|$ is just a convenient code for the probability assigned by Learner outputting forecast p_i to the outcome y_i: it is p_i if $y_i = 1$, and it is $1 - p_i$ if $y_i = 0$). Since $V(\pi^n) = S(\omega^n) \to \infty$ as $n \to \infty$, π is not game-random.

Our argument depends on the superfarthingale V being lower semicomputable. The lower semicomputability of V is easy to check once we finish its definition for the boundary cases. We set $V(p_1, y_1, \ldots, p_n, y_n) := 0$ when $S(y_1, \ldots, y_n) = 0$ (although this case is in fact impossible since S is the universal supermartingale), and we set $V(p_1, y_1, \ldots, p_n, y_n) := \infty$ when $S(y_1, \ldots, y_n) > 0$ but $\prod_{i=1}^{n} |p_i + y_i - 1| = 0$. By the assumption $\forall x \in \Omega^\diamond : \phi(x) \in (0, 1)$ made in the statement of Corollary 1, we have $\mathbb{P}_\phi(\Gamma_{(y_1,\ldots,y_n)}) > 0$. It is an open problem to get rid of this assumption or show that it is essential. (If it turns out to be essential, the next open problem is whether it can be relaxed to the assumption that the set $\phi^{-1}(\{0, 1\}) \subseteq \Omega^\diamond$ is computable.)

6. Prequential probability

In this section we will discuss a question that does not involve the notion of computability but still is closely connected, at least at a philosophical level, to Theorem 1. We have seen that in the prequential framework the notion of randomness splits into game-randomness and measure-randomness. We will see that, in a similar way, the notion of probability splits into game-theoretic probability and measure-theoretic probability. The argument given in the proof of Theorem 1, stating the equivalence of game-randomness and measure-randomness, shows that game-theoretic and measure-theoretic probability coincide on the open sets. The proof of this result will be spelled out explicitly without reference to the proof of Theorem 1: see the proof of Proposition 3 below. This leaves the question of whether game-theoretic and measure-theoretic probability coincide for wider classes of sets. Theorem 2 below states that they indeed coincide on the Borel, and even analytic, sets.

6.1. Measure-theoretic and game-theoretic probability

Let E be a *prequential event*, i.e., a subset of Π. We will define two notions of probability of E, denoted by $\mathbb{P}^{\mathrm{meas}}(E)$ (measure-theoretic) and $\mathbb{P}^{\mathrm{game}}(E)$ (game-theoretic).

Measure-theoretic probability, as formalized by Kolmogorov [16], is standard. In our context, its object of study is the probability measures P on Ω, which we usually represent in the form $P = \mathbb{P}_\phi$ with ϕ a forecasting system. We can apply it to prequential events as follows, in the spirit of [13], Section 10.2. We define the *upper measure-probability* of E as

$$\mathbb{P}^{\mathrm{meas}}(E) := \sup_\phi \mathbb{P}_\phi(E^\phi),$$

where

$$E^\phi := \{\omega \in \Omega \mid \omega^\phi \in E\}.$$

The expression $\mathbb{P}_\phi(E^\phi)$ in the definition of $\mathbb{P}^{\mathrm{meas}}(E)$ is well defined if E is a Borel set: since ω^ϕ is a continuous function of ω, the set $\{\omega \mid \omega^\phi \in E\}$ is also Borel. If E is not Borel, $\mathbb{P}_\phi(E^\phi)$ is understood to be the outer measure of E^ϕ with respect to $\mathbb{P}_\phi$.

The game-theoretic approach to probability is as old as measure-theoretic (see, e.g., [23,31]) but game-theoretic probability was formalized only recently [32,8,29]. Game-theoretic probability can be introduced as either upper or lower probability; in this paper the former is more convenient. The *upper game-probability* of a prequential event E is

$$\mathbb{P}^{\mathrm{game}}(E) := \inf \left\{ \epsilon : \exists V : V(\Lambda) = \epsilon \text{ and } \forall \pi \in E : \limsup_n V(\pi^n) \geq 1 \right\}, \tag{9}$$

where V ranges over the non-negative farthingales. It is clear that nothing changes if lim sup is replaced by sup or lim inf (we can always stop when 1 is reached) and/or if we allow V to range over the non-negative superfarthingales.

Note that upper game-probability is not additive: e.g., both an event and its complement can have upper probability 1. However, it is sub-additive: the upper game-probability of $E_1 \cup E_2$ does not exceed the sum of the upper game-probabilities of E_1 and E_2. The following lemma says that this is also true for countable unions.

Lemma 5. *For any sequence $E_1, E_2, \ldots$ of prequential events,*

$$\mathbb{P}^{\mathrm{game}}\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}^{\mathrm{game}}(E_i).$$

In particular, if $\mathbb{P}^{\mathrm{game}}(E_i) = 0$ for all i, then $\mathbb{P}^{\mathrm{game}}(\bigcup_{i=1}^{\infty} E_i) = 0$.

Proof. It suffices to notice that the sum of a sequence of non-negative farthingales is again a non-negative farthingale. □

Therefore, upper game-probability is an outer measure. The *lower game-probability* of E is defined as

$$\underline{\mathbb{P}}^{\text{game}}(E) := 1 - \mathbb{P}^{\text{game}}(E^c),$$

where E^c is the complement of E. The *exact game-probability* of E exists if $\mathbb{P}^{\text{game}}(E) = \underline{\mathbb{P}}^{\text{game}}(E)$ and is equal to this common value. (Exact game-probability exists for interesting events only in the case of continuous time, as in [33]; in the discrete-time framework of this paper there are many important events for which lower game-probability is close to upper game-probability: see, e.g., [29].) In the rest of this paper we will consider only upper probabilities.

The two notions of probability give two informal notions of randomness that parallel measure-randomness and game-randomness. Namely, Cournot's principle, which is often regarded as the basis for all applications of probability (see, e.g., [28]), can be stated, in our context, as follows: a data sequence is not random if it belongs to a pre-specified event of small upper probability. In the case of game-theoretic probability, a data sequence $\pi \in \Pi$ is regarded as not random if a pre-specified non-negative farthingale starting from 1 becomes large on π. In the case of measure-theoretic probability, a data sequence $\pi \in \Pi$ is regarded as not random if it belongs to a pre-specified critical region E whose probability $\mathbb{P}_\phi(E^\phi)$ is small under any forecasting system ϕ. The coincidence of the two notions of randomness suggests the coincidence of the two notions of probability.

6.2. Coincidence of the two notions of probability

First we prove a simple result showing that upper measure-probability is always less than or equal to upper game-probability.

Proposition 2. *For any set $E \subseteq \Pi$, it is true that $\mathbb{P}^{\text{meas}}(E) \le \mathbb{P}^{\text{game}}(E)$.*

Proof. Fix $E \subseteq \Pi$. It suffices to prove that $\mathbb{P}_\phi(E^\phi) \le U(\Lambda)$ for any forecasting system ϕ and any non-negative farthingale U satisfying $\limsup_n U(\pi^n) \ge 1$ for all $\pi \in E$. Fix such ϕ and U. Then U^ϕ (in the notation of (6)) is a non-negative martingale with respect to ϕ satisfying $\limsup_n U^\phi(\omega^n) \ge 1$ for all $\omega \in E^\phi$. Applying Proposition 1 to $V := U/U(\Lambda)$ and $C := 1/U(\Lambda)$, we can see that indeed $\mathbb{P}_\phi(E^\phi) \le U(\Lambda)$. $\square$

Notice that Proposition 2 holds despite the absence of the requirement of measurability of the (super)farthingale V in (9) (cf. the discussion in [29], pp. 168–169). Even if V is not measurable, V^ϕ is always measurable (any function on $\Omega^\diamond$ is).

A simple modification of the proof of Theorem 1 shows that the following result is true.

Proposition 3. *If $E \subseteq \Pi$ is an open set, $\mathbb{P}^{\text{meas}}(E) = \mathbb{P}^{\text{game}}(E)$.*

Proof. Fix an open set $E \subseteq \Pi$. The idea of the proof is to construct the smallest non-negative superfarthingale (denoted by W below) such that $\limsup_n W(\pi^n) \ge 1$ for all $\pi \in E$.

Represent E as the union $E = \bigcup_{i=1}^\infty E_i$ of a nested sequence $E_1 \subseteq E_2 \subseteq \cdots$ of open sets such that each E_i satisfies

$$\forall \pi \in \Pi : \pi \in E_i \Longrightarrow \Gamma_{\pi^i} \subseteq E_i \tag{10}$$

(informally, E_i is a property of the first i forecasts and outcomes). For each $i = 1, 2, \ldots$, define a superfarthingale W_i as follows. For all $x \in \Pi^{\ge i}$, set

$$W_i(x) := \begin{cases} 1 & \text{if } \Gamma_x \subseteq E_i \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The rest of the definition is inductive. If $W_i(x)$ is already defined for $x \in \Pi^n$, $n = i, i-1, \ldots, 1$, define $W_i(x)$, for each $x \in \Pi^{n-1}$, by (7).

To prove the inequality $\mathbb{P}^{\text{meas}}(E) \ge \mathbb{P}^{\text{game}}(E)$ (this is all we need to do, in view of Proposition 2), let us first check that $W_i(\Lambda) \le \mathbb{P}^{\text{meas}}(E)$. Suppose that, on the contrary, $W_i(\Lambda) > \mathbb{P}^{\text{meas}}(E)$. Construct a forecasting system ϕ as follows. For each $x \in \Omega^n$, $n = 0, 1, \ldots, i-1$, choose $\phi(x)$ such that

$$\begin{aligned}&(1-\phi(x))W_i(x^\phi, \phi(x), 0) + \phi(x) W_i(x^\phi, \phi(x), 1)\\ &\quad\ge \sup_{p\in[0,1]} \big((1-p)W_i(x^\phi, p, 0) + pW_i(x^\phi, p, 1)\big) - \epsilon/i = W_i(x^\phi) - \epsilon/i,\end{aligned}$$

where $\epsilon > 0$ satisfies $W_i(\Lambda) > \mathbb{P}^{\text{meas}}(E) + \epsilon$. For each $x \in \Omega^{\ge i}$, set, e.g., $\phi(x) := 0$. Since the function $x \in \Omega^\diamond \mapsto S(x) := W_i^\phi(x) + \epsilon n/i$, where n is the length of x, is a submartingale with respect to ϕ, we have

$$\begin{aligned}\mathbb{P}^{\text{meas}}(E) &\ge \mathbb{P}_\phi(E^\phi) \ge \mathbb{P}_\phi(E_i^\phi)\\ &= \mathbb{P}_\phi\left\{\omega \mid W_i^\phi(\omega^i) = 1\right\} = \mathbb{E}_\phi W_i^\phi(\omega^i) = \mathbb{E}_\phi(S(\omega^i) - \epsilon)\\ &\ge S(\Lambda) - \epsilon = W_i^\phi(\Lambda) - \epsilon = W_i(\Lambda) - \epsilon > \mathbb{P}^{\text{meas}}(E),\end{aligned}$$

a contradiction.

Setting $W := \sup_i W_i$, we obtain a non-negative superfarthingale satisfying $W(\Lambda) \le \mathbb{P}^{\text{meas}}(E)$ and $\limsup_n W(\pi^n) \ge 1$ for all $\pi \in E$. Therefore, $\mathbb{P}^{\text{game}}(E) \le \mathbb{P}^{\text{meas}}(E)$. $\square$

It turns out that the two notions of prequential probability coincide on all analytic sets (this is a wide class containing, e.g., all Borel sets; a precise definition can be found in, e.g., [14]).

Theorem 2. *If $E \subseteq \Pi$ is an analytic set, $\mathbb{P}^{\mathrm{meas}}(E) = \mathbb{P}^{\mathrm{game}}(E)$.*

This is the second main result of this paper. We will prove it in the next section. This result contains Proposition 3 as a special case. We give an independent proof of Proposition 3 since it is much more constructive than the proof of the general statement.

7. Proof of Theorem 2

The inequality $\leq$ in Theorem 2 immediately follows from Proposition 2, and so it suffices to prove the inequality $\geq$.

We start from proving a special case of Theorem 2.

Lemma 6. *If $E \subseteq \Pi$ is a compact set, $\mathbb{P}^{\mathrm{meas}}(E) = \mathbb{P}^{\mathrm{game}}(E)$.*

Proof. Fix a compact prequential event $E \subseteq \Pi$. (Of course, "compact" is the same thing as "closed" in this context.) Represent E as the intersection $E = \bigcap_{i=1}^{\infty} E_i$ of a nested sequence $E_1 \supseteq E_2 \supseteq \cdots$ of closed sets such that (10) is satisfied for all i. For each $i = 1, 2, \ldots$, define a superfarthingale W_i by setting (11) for all $x \in \Pi^{\geq i}$ and then proceeding inductively as follows. If $W_i(x)$ is already defined for $x \in \Pi^n$, $n = i, i-1, \ldots, 1$, define $W_i(x)$, for each $x \in \Pi^{n-1}$, by (7). It is clear that $W_1 \geq W_2 \geq \cdots$.

Let us check that $W_i(x)$ is upper semicontinuous as a function of $x \in \Pi^{\diamond}$. By (11) this is true for $x \in \Pi^{\geq i}$. Suppose this is true for $x \in \Pi^n$, $n \in \{i, i-1, \ldots, 2\}$, and let us prove that it is true for $x \in \Pi^{n-1}$, using the inductive definition (7). It is clear that $f(x,p) := (1-p)W_i(x,p,0) + pW_i(x,p,1)$ is upper semicontinuous as function of $p \in [0,1]$ and $x \in \Pi^{n-1}$. It is well known that $\sup_p f(x,p)$ is upper semicontinuous whenever f is upper semicontinuous and x and p range over compact sets (see, e.g., [9], Theorem I.2(d); a simple proof of a slightly more general fact will be given below in Lemma 7). Therefore, $W_i(x) = \sup_{p\in[0,1]} f(x,p)$ is an upper semicontinuous function of $x \in \Pi^{n-1}$.

An important implication of the upper semicontinuity of W_i and the compactness of $[0,1]$ is that the supremum in (7) is attained: it is easy to check that an upper semicontinuous function attains its supremum over a compact set (cf. [10], Problem 3.12.23(g)). For each $i = 1, 2, \ldots$, we can now define a forecasting system ϕ_i as follows. For each $x \in \Omega^n$, $n = 0, 1, \ldots, i-1$, choose $\phi_i(x)$ such that

$$(1-\phi_i(x))W_i(x^{\phi_i}, \phi_i(x), 0) + \phi_i(x)W_i(x^{\phi_i}, \phi_i(x), 1) = \sup_p \big((1-p)W_i(x^{\phi_i}, p, 0) + pW_i(x^{\phi_i}, p, 1)\big) = W_i(x^{\phi_i})$$

(this is an inductive definition; in particular, x^{ϕ_i} is already defined at the time of defining $\phi_i(x)$). For $x \in \Omega^{\geq i}$, set, for example, $\phi_i(x) := 0$. Since $W_i^{\phi_i}$ is a martingale with respect to ϕ_i, we have $\mathbb{P}_{\phi_i}(E_i^{\phi_i}) = W_i(\Lambda)$.

Since the set Φ of all forecasting systems is compact in the product topology, the sequence ϕ_i has a convergent subsequence ϕ_{i_k}, $k = 1, 2, \ldots$; let $\phi := \lim_{k\to\infty} \phi_{i_k}$. We assume, without loss of generality, $i_1 < i_2 < \cdots$. Set

$$c := \inf_i W_i(\Lambda) = \lim_{i\to\infty} W_i(\Lambda).$$

Fix an arbitrarily small $\epsilon > 0$. Let us prove that $\mathbb{P}_{\phi}(E^{\phi}) \geq c - \epsilon$. Let $K \in \mathbb{N}$. The restriction of $\mathbb{P}_{\phi_{i_k}}$ to Ω^{i_K} (more formally, the probability measure assigning weight $\mathbb{P}_{\phi_{i_k}}(\Gamma_x)$ to each singleton $\{x\}$, $x \in \Omega^{i_K}$) comes within ϵ of the restriction of $\mathbb{P}_{\phi}$ to Ω^{i_K} in total variation distance from some k on; let the total variation distance be at most ϵ for all $k \geq K' \geq K$. Let $k \geq K'$. Since $\mathbb{P}_{\phi_{i_k}}(E_{i_k}^{\phi_{i_k}}) \geq c$, it is also true that $\mathbb{P}_{\phi_{i_k}}(E_{i_K}^{\phi_{i_k}}) \geq c$; therefore, it is true that $\mathbb{P}_{\phi}(E_{i_K}^{\phi_{i_k}}) \geq c - \epsilon$. By Fatou's lemma, we now obtain

$$\mathbb{P}_{\phi}\left(\limsup_k E_{i_K}^{\phi_{i_k}}\right) \geq \limsup_{k\to\infty} \mathbb{P}_{\phi}(E_{i_K}^{\phi_{i_k}}) \geq c - \epsilon. \tag{12}$$

Let us check that

$$\limsup_k E_{i_K}^{\phi_{i_k}} \subseteq E_{i_K}^{\phi}. \tag{13}$$

Indeed, let $\omega \notin E_{i_K}^{\phi}$, i.e., $\omega^{\phi} \notin E_{i_K}$. Since $\phi_{i_k} \to \phi$ in the product topology and the set E_{i_K} is closed, $\omega^{\phi_{i_k}} \notin E_{i_K}$ from some k on. This means that $\omega \in E_{i_K}^{\phi_{i_k}}$ for only finitely many k, i.e., $\omega \notin \limsup_k E_{i_K}^{\phi_{i_k}}$.

From (12) and (13) we can see that $\mathbb{P}_{\phi}(E_{i_K}^{\phi}) \geq c - \epsilon$, for all $K \in \mathbb{N}$. This implies $\mathbb{P}_{\phi}(E^{\phi}) \geq c - \epsilon$. Since this holds for all ϵ, $\mathbb{P}_{\phi}(E^{\phi}) \geq c$.

The rest of the proof is easy: since

$$\mathbb{P}^{\mathrm{game}}(E) \leq c \leq \mathbb{P}_{\phi}(E^{\phi}) \leq \mathbb{P}^{\mathrm{meas}}(E) \leq \mathbb{P}^{\mathrm{game}}(E)$$

(the last inequality following from Proposition 2), we have

$$\mathbb{P}^{\mathrm{game}}(E) = c = \mathbb{P}_{\phi}(E^{\phi}) = \mathbb{P}^{\mathrm{meas}}(E). \quad \square$$

In the proof of Lemma 6 we referred to the following analogue of Lemma 4.

Lemma 7. *Suppose X and Y are topological spaces and Y is compact. If a function $f : X \times Y \to \mathbb{R}$ is upper semicontinuous, then the function $x \in X \mapsto g(x) := \sup_{y\in Y} f(x,y)$ is also upper semicontinuous.*

Proof. For any $c \in \mathbb{R}$, we are required to show that the set $G := \{x \mid \sup_y f(x, y) < c\}$ is open. Let $x \in G$. For any $y \in Y$ there exists a neighbourhood O'_y of x and a neighbourhood O''_y of y such that, for some $\epsilon > 0$, $f(x', y') < c - \epsilon$ for all $x' \in O'_y$ and all $y' \in O''_y$. By the compactness of Y, there is a finite family $O''_{y_1}, \dots, O''_{y_K}$ that covers Y. The intersection of $O'_{y_1}, \dots, O'_{y_K}$ will contain x and will be a subset of G. Therefore, G is indeed open. □

The argument in [9], proof of Theorem I.2(d), is even simpler, but it assumes that X is compact (which is, however, sufficient for the purpose of Lemma 6).

The idea of the proof of Theorem 2 is to extend Lemma 6 to the analytic sets using Choquet's capacitability theorem (stated below). Remember that a function γ (such as $\mathbb{P}^{\text{game}}$ or $\mathbb{P}^{\text{meas}}$) mapping the power set of a topological space X (such as Π) to $[0, \infty)$ is a *capacity* (on X) if:

- for any subsets A and B of X,

$$A \subseteq B \Longrightarrow \gamma(A) \le \gamma(B); \tag{14}$$

- for any nested increasing sequence $A_1 \subseteq A_2 \subseteq \cdots$ of arbitrary subsets of X,

$$\gamma\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \gamma(A_i); \tag{15}$$

- for any nested decreasing sequence $K_1 \supseteq K_2 \supseteq \cdots$ of compact sets in X,

$$\gamma\left(\bigcap_{i=1}^{\infty} K_i\right) = \lim_{i\to\infty} \gamma(K_i). \tag{16}$$

Condition (16) is sometimes replaced by a different condition which is equivalent to (16) for compact metrizable spaces X: cf. [14], Definition 30.1.

It turns out that both $\mathbb{P}^{\text{game}}$ and $\mathbb{P}^{\text{meas}}$ are capacities. We start from $\mathbb{P}^{\text{game}}$.

Theorem 3. *The set function $\mathbb{P}^{\text{game}}$ is a capacity.*

It is obvious that $\mathbb{P}^{\text{game}}$ satisfies condition (14). The following two statements establish conditions (15) and (16). Condition (16) is easier to check: it can be extracted from the proof of Lemma 6.

Lemma 8. *If $K_1 \supseteq K_2 \supseteq \cdots$ is a nested sequence of compact sets in Π,*

$$\mathbb{P}^{\text{game}}\left(\bigcap_{i=1}^{\infty} K_i\right) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(K_i). \tag{17}$$

Proof. We will use the equality $\mathbb{P}^{\text{game}}(E) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(E_i)$, in the notation of the proof of Lemma 6. This equality follows from

$$\mathbb{P}^{\text{game}}(E) = c = \lim_{i\to\infty} W_i(\Lambda) \ge \lim_{i\to\infty} \mathbb{P}^{\text{game}}(E_i)$$

(the opposite inequality is obvious).

Represent each K_n in the form $K_n = \bigcap_{i=1}^{\infty} E_i$ where $E_1 \supseteq E_2 \supseteq \cdots$ and each E_i satisfies (10); we will write $K_{n,i}$ in place of E_i. Without loss of generality we will assume that $K_{1,i} \supseteq K_{2,i} \supseteq \cdots$ for all i. Then the set $K := \bigcap_{i=1}^{\infty} K_i$ can be represented as $K = \bigcap_{i=1}^{\infty} K_{i,i}$, and so (17) follows from

$$\begin{aligned}\mathbb{P}^{\text{game}}(K) &= \mathbb{P}^{\text{game}}\left(\bigcap_{i=1}^{\infty} K_{i,i}\right) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(K_{i,i}) = \lim_{n\to\infty}\lim_{i\to\infty} \mathbb{P}^{\text{game}}(K_{n,i}) \\ &= \lim_{n\to\infty} \mathbb{P}^{\text{game}}\left(\bigcap_{i=1}^{\infty} K_{n,i}\right) = \lim_{n\to\infty} \mathbb{P}^{\text{game}}(K_n). \quad \square\end{aligned}$$

To check condition (15) for $\mathbb{P}^{\text{game}}$, we will need the game-theoretic version, proved in [30], of Lévy's zero-one law ([20], Section 41). For each $x \in \Pi^{\diamond}$, define the *conditional upper game-probability* of $E \subseteq \Pi$ by

$$\mathbb{P}^{\text{game}}(E \mid x) := \inf\left\{\epsilon \mid \exists V : V(x) = \epsilon \text{ and } \forall \pi \in E \cap \Gamma_x : \limsup_n V(\pi^n) \ge 1\right\},$$

where V ranges over the non-negative (super)farthingales.

Proposition 4 ([30]). *Let $E \subseteq \Pi$. For almost all $\pi \in E$,*

$$\mathbb{P}^{\text{game}}(E \mid \pi^n) \to 1 \tag{18}$$

as $n \to \infty$. (In the sense that there exists a prequential event N such that $\mathbb{P}^{\text{game}}(N) = 0$ and (18) holds for all $\pi \in E \setminus N$.)

Proof. Without loss of generality we replace (18) by

$$\liminf_{n\to\infty} \mathbb{P}^{\text{game}}(E \mid \pi^n) \geq a, \tag{19}$$

where $a \in (0, 1)$ is a given rational number. (Indeed, by Lemma 5, if (19) holds for almost all $\pi \in E$ for each rational $a \in (0, 1)$, the intersection of (19) over all such a also holds for almost all $\pi \in E$, and so (18) holds for almost all $\pi \in E$ as well.) It suffices to construct a non-negative farthingale V starting from 1 that tends to ∞ on the sequences $\pi \in E$ for which (19) is not true.

Let π be any sequence in Π; we will define $V(\pi^n)$ by induction for $n = 1, 2, \ldots$ (intuitively, we will describe a gambling strategy with capital process V). Start with 1 monetary unit: $V(\Lambda) := 1$. Keep setting $V(\pi^n) := 1$, $n = 1, 2, \ldots$, until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$ (if this never happens, $V(\pi^n)$ will be 1 for all n). Let N_1 be the first n when this happens: $\mathbb{P}^{\text{game}}(E \mid \pi^{N_1}) < a$ but $\mathbb{P}^{\text{game}}(E \mid \pi^n) \geq a$ for all $n < N_1$. Choose a non-negative farthingale S_1 starting at π^{N_1} from 1, $S_1(\pi^{N_1}) = 1$, whose upper limit exceeds $1/a$ on all extensions of π^{N_1} in E. Keep setting $V(\pi^n) := S_1(\pi^n)$, $n = N_1, N_1 + 1, \ldots$, until $S_1(\pi^n)$ reaches a value $s_1 > 1/a$. After that keep setting $V(\pi^n) := V(\pi^{n-1})$ until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$. Let N_2 be the first n when this happens. Choose a non-negative farthingale S_2 starting at π^{N_2} from s_1, $S_2(\pi^{N_2}) = s_1$, whose upper limit exceeds s_1/a on all extensions of π^{N_2} in E. Keep setting $V(\pi^n) := S_2(\pi^n)$, $n = N_2, N_2 + 1, \ldots$, until $S_2(\pi^n)$ reaches a value $s_2 > s_1(1/a) > (1/a)^2$. After that keep setting $V(\pi^n) := V(\pi^{n-1})$ until $\mathbb{P}^{\text{game}}(E \mid \pi^n) < a$. Let N_3 be the first n when this happens. Choose a non-negative farthingale S_3 starting at π^{N_3} from s_2 whose upper limit exceeds s_2/a on all extensions of π^{N_3} in E. Keep setting $V(\pi^n) := S_3(\pi^n)$, $n = N_3, N_3 + 1, \ldots$, until S_3 reaches a value $s_3 > s_2(1/a) > (1/a)^3$. And so on. □

Lemma 9. *If $A_1 \subseteq A_2 \subseteq \cdots \subseteq \Pi$ is a nested sequence of prequential events,*

$$\mathbb{P}^{\text{game}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(A_i). \tag{20}$$

Proof. Let $A_1, A_2, \ldots$ be a nested increasing sequence of prequential events. The non-trivial inequality in (20) is $\leq$. For each A_i the process

$$S_i(x) := \mathbb{P}^{\text{game}}(A_i \mid x)$$

is a non-negative superfarthingale (see Lemma 10 below). By Proposition 4, $\limsup_n S_i(\pi^n) \geq 1$ for almost all $\pi \in A_i$. The sequence S_i is increasing, $S_1 \leq S_2 \leq \cdots$, so the limit $S := \lim_{i\to\infty} S_i = \sup_i S_i$ exists and is a non-negative superfarthingale such that $S(\Lambda) = \lim_{i\to\infty} \mathbb{P}^{\text{game}}(A_i)$ and $\limsup_n S(\pi^n) \geq 1$ for almost all $\pi \in \bigcup_i A_i$ (by Lemma 5). We can get rid of "almost" by adding to S a non-negative farthingale V that starts at $V(\Lambda) < \epsilon$, for an arbitrarily small $\epsilon > 0$, and satisfies $\limsup_n V(\pi^n) \geq 1$ for all $\pi \in \bigcup_i A_i$ violating $\limsup_n S(\pi^n) \geq 1$. □

Lemma 10. *For any prequential event E, the function $x \in \Pi^\diamond \mapsto \mathbb{P}^{\text{game}}(E \mid x)$ is a superfarthingale.*

Proof. Suppose there are $x \in \Pi^\diamond$ and $p \in [0, 1]$ such that

$$\mathbb{P}^{\text{game}}(E \mid x) < (1-p)\mathbb{P}^{\text{game}}(E \mid x, p, 0) + p\mathbb{P}^{\text{game}}(E \mid x, p, 1).$$

Then there exists a non-negative farthingale V with $\limsup_n V(\pi^n) \geq 1$ for all $\pi \in E \cap \Gamma_x$ that satisfies

$$V(x) < (1-p)\mathbb{P}^{\text{game}}(E \mid x, p, 0) + p\mathbb{P}^{\text{game}}(E \mid x, p, 1)$$

and, therefore,

$$(1-p)V(x, p, 0) + pV(x, p, 1) < (1-p)\mathbb{P}^{\text{game}}(E \mid x, p, 0) + p\mathbb{P}^{\text{game}}(E \mid x, p, 1).$$

The last inequality implies that there exists $j \in \{0, 1\}$ such that $V(x, p, j) < \mathbb{P}^{\text{game}}(E \mid x, p, j)$, which is impossible. □

This completes the proof of Theorem 3. Let us now check that upper measure-probability is also a capacity.

Lemma 11. *The set function $\mathbb{P}^{\text{meas}}$ is a capacity.*

Proof. Property (14) is obvious for $\mathbb{P}^{\text{meas}}$. Property (16) follows from Lemmas 6 and 8.

Let us now check the remaining property (15), with $\mathbb{P}^{\text{meas}}$ as γ. Suppose there exists an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq X$ of prequential events such that

$$\mathbb{P}^{\text{meas}}\left(\bigcup_{i=1}^{\infty} A_i\right) > \lim_{i\to\infty} \mathbb{P}^{\text{meas}}(A_i).$$

Let ϕ be a forecasting system satisfying

$$\mathbb{P}_\phi\left(\bigcup_{i=1}^{\infty} A_i^\phi\right) > \lim_{i\to\infty} \mathbb{P}^{\mathrm{meas}}(A_i).$$

Then ϕ will satisfy $\mathbb{P}_\phi\left(\bigcup_{i=1}^{\infty} A_i^\phi\right) > \lim_{i\to\infty} \mathbb{P}_\phi(A_i^\phi)$, which contradicts $\mathbb{P}_\phi$ being a capacity (see, e.g., [14], Exercise 30.3). $\square$

In combination with Choquet's capacitability theorem, Theorem 3 and Lemma 11 allow us to finish the proof of Theorem 2.

Theorem 4 (*Choquet's Capacitability Theorem, [3]*). *If X is a compact metrizable space, γ is a capacity on X, and $E \subseteq X$ is an analytic set,*

$$\gamma(E) = \sup\{\gamma(K) \mid K \text{ is compact}, K \subseteq E\}.$$

For a proof of Choquet's theorem, see, e.g., [14], Theorem 30.13.

Proof of Theorem 2. Combining Choquet's capacitability theorem (applied to the compact metrizable space Π), Lemma 6, Theorem 3, and Lemma 11, we obtain

$$\mathbb{P}^{\mathrm{meas}}(E) = \sup_{K\subseteq E} \mathbb{P}^{\mathrm{meas}}(K) = \sup_{K\subseteq E} \mathbb{P}^{\mathrm{game}}(K) = \mathbb{P}^{\mathrm{game}}(E),$$

K ranging over the compact sets in Π. $\square$

Remark 5. The fact that upper game-probability and upper measure-probability are capacities has allowed us to prove their coincidence on the analytic sets, and it might be useful for other purposes as well. In general, neither of these capacities is *strongly subadditive*, in the sense of satisfying

$$\gamma(A\cup B) + \gamma(A\cap B) \le \gamma(A) + \gamma(B)$$

for all prequential events A and B. To demonstrate this it suffices, in view of Theorem 2, to find analytic sets A and B that violate

$$\mathbb{P}^{\mathrm{game}}(A\cup B) + \mathbb{P}^{\mathrm{game}}(A\cap B) \le \mathbb{P}^{\mathrm{game}}(A) + \mathbb{P}^{\mathrm{game}}(B). \tag{21}$$

We can define $\mathbb{P}^{\mathrm{game}}(E)$ for subsets E of Π^n by (9) with $\limsup_n$ omitted. This is an example of subsets A and B of Π^2 for which (21) is violated:

$$A = \left\{\left(0, 0, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, 0, 0\right)\right\}, \tag{22}$$

$$B = \left\{\left(0, 0, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 1, 0, 0\right)\right\}. \tag{23}$$

For these subsets we have

$$\mathbb{P}^{\mathrm{game}}(A\cup B) + \mathbb{P}^{\mathrm{game}}(A\cap B) = 1 + \frac{1}{2} > \frac{1}{2} + \frac{1}{2} = \mathbb{P}^{\mathrm{game}}(A) + \mathbb{P}^{\mathrm{game}}(B).$$

To obtain an example of subsets A and B of the full prequential space Π for which (21) is violated, it suffices to add $00\ldots$ at the end of each element of the sets A and B defined by (22) and (23).

Acknowledgements

We are grateful to A. Philip Dawid and Glenn Shafer for illuminating discussions and to the referees of the conference and journal versions of this paper for useful comments. First author was supported by EPSRC (grant EP/F002998/1). Second author was supported by ANR (grant NAFIT ANR-08-EMER-008-01).

Appendix. Effective topology

In this section we will give definitions of various notions connected with computability in topological spaces, mainly following Martin-Löf [22] (see also [12], Appendix C.2). The details of the definitions become important only in the proofs. We will use the terminology of Engelking [10].

An *effective topological space* is a second-countable topological space with a fixed numbering $(U_k)_{k=1}^\infty$ of its countable base. In other words, an effective topological space is a triple $(X, \mathcal{O}, (U_k)_{k=1}^\infty)$, where $(X, \mathcal{O})$ is a topological space and $(U_k)_{k=1}^\infty$ is a numbering of its countable base. The family $(U_k)_{k=1}^\infty$ is called the *effective base* of the effective topological space, and its elements are called *basic sets*. Finite unions of basic sets are called *simple sets*. We do not distinguish between two effective topological spaces $(X, \mathcal{O}, (U_k)_{k=1}^\infty)$ and $(X', \mathcal{O}', (U'_k)_{k=1}^\infty)$ if $(X, \mathcal{O}) = (X', \mathcal{O}')$ and there exists a computable bijection $f : \mathbb{N} \to \mathbb{N}$ such that $U'_k = U_{f(k)}$ for all k.

Example 1 ($\mathbb{N}$). The usual discrete topology on $\mathbb{N}$ has as its base the set of all singletons $\{k\}$, $k \in \mathbb{N}$. They can serve as the effective base, $U_k := \{k\}$.

Example 2 ($\mathbb{R}$). The topology on $\mathbb{R}$ has as its base the set of all intervals (a, b), $a < b$. To make $\mathbb{R}$ into an effective topological space, fix a computable enumeration (a_k, b_k), $k = 1, 2, \ldots$, of all intervals with rational end-points, and take $U_k := (a_k, b_k)$ as the effective base.

Example 3 (Ω). The topology on $\Omega := \{0, 1\}^\infty$ is the usual product topology, which makes Ω a compact topological space. To make it into an effective topological space, fix a computable bijection $f : \mathbb{N} \to \Omega^\diamond$ and take $U_k := \Gamma_{f(k)}$ as the effective base.

Example 4 (Φ). The basic sets in Φ (the set of all forecasting systems) have the form

$$\left\{\phi \in \Phi \mid a(x) < \phi(x) < b(x), \forall x \in \Omega^{\le n}\right\} \tag{24}$$

for some $n \in \mathbb{N}$ and $a, b : \Omega^{\le n} \to \mathbb{Q}$. Let (n_k, a_k, b_k), $k = 1, 2, \ldots$, be a computable enumeration of all such triples (n, a, b). Set U_k to (24) with $(n, a, b) := (n_k, a_k, b_k)$.

Example 5 (Π). The topology on the prequential space Π is the standard product topology of $[0, 1] \times \{0, 1\} \times [0, 1] \times \{0, 1\} \times \cdots$. The basic sets are

$$\left\{(p_1, y_1, p_2, y_2, \ldots) \in \Pi \mid a_1 < p_1 < b_1, y_1 = c_1, \ldots, a_n < p_n < b_n, y_n = c_n\right\} \tag{25}$$

where n ranges over $\mathbb{N}$, $a_i, b_i \in \mathbb{Q}$, and $c_i \in \{0, 1\}$, $i = 1, \ldots, n$. Let

$$(n_k, a_{1,k}, b_{1,k}, c_{1,k}, \ldots, a_{n_k,k}, b_{n_k,k}, c_{n_k,k}) \tag{26}$$

be a computable enumeration of all such sequences $(n, a_1, b_1, c_1, \ldots, a_n, b_n, c_n)$. We can define U_k as (25) with (26) in place of $(n, a_1, b_1, c_1, \ldots, a_n, b_n, c_n)$.

Let X' and X'' be two effective topological spaces with effective bases $(U'_k)_{k=1}^\infty$ and $(U''_k)_{k=1}^\infty$, respectively. The Cartesian product of X' and X'' is the product of the topological spaces X' and X'' equipped with the effective base $(U_k)_{k=1}^\infty$, where $U_{f(k',k'')} := U'_{k'} \times U''_{k''}$ and $f : \mathbb{N}^2 \to \mathbb{N}$ is a fixed computable bijection. We will be particularly interested in the product $\Omega \times \Phi$; sometimes we will need products of more than two spaces, such as $\Omega \times \Phi \times \mathbb{R} := (\Omega \times \Phi) \times \mathbb{R}$.

Let X be an effective topological space with effective base $(U_k)_{k=1}^\infty$. As described in the previous paragraph, we define the structure of an effective topological space on the power set X^n, $n \in \mathbb{N}$; let the effective base in X^n be $(U_k^{(n)})_{k=1}^\infty$. For $n = 0$, X^0 is the trivial one-element effective topological space with all $U_k = X^0$, $k \in \mathbb{N}$. The set X^* of all finite sequences of elements of X is equipped with the topology of the direct sum of X^n, $n \ge 0$. An effective base in it can be defined by $U_{f(n,k)} := U_k^{(n)}$, where $f : (\mathbb{N} \cup \{0\}) \times \mathbb{N} \to \mathbb{N}$ is a computable bijection.

Let X be a fixed effective topological space with effective base $(U_k)_{k=1}^\infty$. An open set $G \subseteq X$ is said to be *effectively open* if it can be represented in the form $G = \cup\{U_k \mid k \in A\}$ for a recursively enumerable set $A \subseteq \mathbb{N}$. In the main part of this paper, for any effectively open set G we restrict ourselves only to its representations $\cup\{U_k \mid k \in A\}$ such that

$$\overline{U_k} \subseteq G; \tag{27}$$

this can be done without loss of generality for all specific effective topological spaces that we need. A *computable sequence of open sets* is a sequence of open sets $G_1, G_2, \ldots$ such that there exists a recursively enumerable set $A \subseteq \mathbb{N}^2$ satisfying $G_m = \cup\{U_k \mid (m, k) \in A\}$ for all $m \in \mathbb{N}$. A *computable family of sequences of open sets* is a family $(G_{l,m})$, $l, m \in \mathbb{N}$, of sequences of open sets such that there exists a recursively enumerable set $A \subseteq \mathbb{N}^3$ satisfying $G_{l,m} = \cup\{U_k \mid (l, m, k) \in A\}$ for all l, m. The existence of a universal Turing machine immediately implies

Lemma 12. *There exists a computable family $(G_{l,m})$ of sequences of open sets such that for any computable sequence G'_m of open sets there exists $l \in \mathbb{N}$ such that $G'_m = G_{l,m}$ for all $m \in \mathbb{N}$.*

Any computable family of sequences of open sets satisfying the condition in Lemma 12 will be called a *universal computable family of sequences of open sets*.

A function $f : X \to \mathbb{R} \cup \{\infty\}$ is called *lower semicomputable* if the set $\{(x, r) \mid x \in X, r \in \mathbb{R}, f(x) > r\}$ is effectively open in $X \times \mathbb{R}$. Similarly, a function $f : X \to \overline{\mathbb{N}}_0$ is *lower semicomputable* if the set $\{(x, r) \mid x \in X, r \in \mathbb{N}, f(x) \ge r\}$ is effectively open in $X \times \mathbb{N}$. A sequence $f_1, f_2, \ldots$ of lower semicomputable functions $f_l : X \to \mathbb{R} \cup \{\infty\}$ is called *computable* if the set $\{(l, x, r) \mid x \in X, r \in \mathbb{R}, f_l(x) > r\}$ is effectively open in $\mathbb{N} \times X \times \mathbb{R}$. The existence of a universal Turing machine also implies

Lemma 13. *There exists a computable sequence $f_1, f_2, \ldots$ of lower semicomputable functions $f_l : X \to [0, \infty]$ that contains every lower semicomputable function $f : X \to [0, \infty]$.*

Any computable sequence of lower semicomputable functions satisfying the condition in Lemma 13 will be called a *universal computable sequence of lower semicomputable functions*.

It is not difficult to check that the notion of lower semicomputability is an effective version of the standard topological notion of lower semicontinuity: a function $f : X \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous (in the usual sense of $\{x \in X \mid f(x) > r\}$ being open for each $r \in \mathbb{R}$) if and only if the set $\{(x, r) \mid x \in X, r \in \mathbb{R}, f(x) > r\}$ is open in the product $X \times \mathbb{R}$.

Lemma 14. *If $f : X \to [0, \infty]$ is lower semicomputable and $C \in \mathbb{N} \cup \{0\}$, the set $\{x \mid f(x) > C\}$ is effectively open.*

Proof. Let $\cup\{U_k \mid k \in A\}$, with $A \subseteq \mathbb{N}$ recursively enumerable, be a representation of the effectively open set $\{(x, r) \mid x \in X, r \in \mathbb{R}, f(x) > r\}$ as a union of basic sets in $X \times \mathbb{R}$. The set $\{x \mid f(x) > C\}$ can be represented as the union of the basic sets $\{x \in X \mid \exists r \in \mathbb{R} : (x, r) \in U_k\}$ over $k \in A$ such that $\sup\{r \mid \exists x : (x, r) \in U_k\} > C$. $\square$

A function $f : X \to \mathbb{R}$ is called *computable* if both f and $-f$ are lower semicomputable. It is easy to see that the analogue of Lemma 13 does not hold for computable functions.

References

[1] Laurent Bienvenu, Caractérisations de l'aléatoire par les jeux: impredictibilité et stochasticité, Ph.D. Thesis, Université de Provence, 2008.
[2] Alexey Chernov, Alexander Shen, Nikolai Vereshchagin, Vladimir Vovk, On-line probability, complexity and randomness, in: Yoav Freund, László Györfi, György Turán, Thomas Zeugmann (Eds.), Proceedings of the Nineteenth International Conference on Algorithmic Learning Theory, in: Lecture Notes in Artificial Intelligence, vol. 5254, Springer, Berlin, 2008, pp. 138–153.
[3] Gustave Choquet, Theory of capacities, Annales de l'institut Fourier 5 (1954) 131–295.
[4] Thomas M. Cover, Joy A. Thomas, Elements of Information Theory, second edition, Wiley, Hoboken, NJ, 2006.
[5] A. Philip Dawid, Statistical theory: the prequential approach (with discussion), Journal of the Royal Statistical Society A 147 (1984) 278–292.
[6] A. Philip Dawid, Calibration-based empirical probability (with discussion), Annals of Statistics 13 (1985) 1251–1285.
[7] A. Philip Dawid, Prequential analysis, in: Samuel Kotz, Campbell B. Read, David L. Banks (Eds.), Encyclopedia of Statistical Sciences, vol. 1 (update), Wiley, New York, 1997, pp. 464–470.
[8] A. Philip Dawid, Vladimir Vovk, Prequential probability: principles and properties, Bernoulli 5 (1999) 125–162.
[9] Claude Dellacherie, Ensembles analytiques, capacités, mesures de Hausdorff, in: Lecture Notes in Mathematics, vol. 295, Springer, Berlin, 1972.
[10] Ryszard Engelking, General Topology, second edition, Heldermann, Berlin, 1989.
[11] Peter Gács, Exact expressions for some randomness tests, Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 26 (1980) 385–394.
[12] Peter Gács, Uniform test of algorithmic randomness over a general space, Theoretical Computer Science 341 (2005) 91–137.
[13] Peter J. Huber, Robust Statistics, Wiley, New York, 1981.
[14] Alexander S. Kechris, Classical Descriptive Set Theory, Springer, New York, 1995.
[15] John L. Kelly, A new interpretation of information rate, Bell System Technical Journal 35 (1956) 917–926.
[16] Andrei N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin, 1933, English translation: Foundations of the Theory of Probability, Chelsea, New York, 1950.
[17] Andrei N. Kolmogorov, Vladimir A. Uspensky, Algorithms and randomness, Theory of Probability and its Applications 32 (1987) 389–412.
[18] Leonid A. Levin, On the notion of a random sequence, Soviet Mathematics Doklady 14 (1973) 1413–1416.
[19] Leonid A. Levin, Uniform tests of randomness, Soviet Mathematics Doklady 17 (1976) 337–340.
[20] Paul Lévy, Théorie de l'addition des variables aléatoires, second edition: 1954, Gauthier-Villars, Paris, 1937.
[21] Per Martin-Löf, The definition of random sequences, Information and Control 9 (1966) 602–619.
[22] Per Martin-Löf, Notes on Constructive Mathematics, Almqvist & Wiksell, Stockholm, 1970.
[23] Richard von Mises, Grundlagen der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift 5 (1919) 52–99.
[24] Judea Pearl, Causal diagrams for empirical research (with discussion), Biometrika 82 (1995) 669–710.
[25] Judea Pearl, Causation, action and counterfactuals, in: Alex Gammerman (Ed.), Computational Learning and Probabilistic Reasoning, Wiley, Chichester, 1996, pp. 103–124 (chapter 6).
[26] Jan Reimann, Effectively closed classes of measures and randomness, Annals of Pure and Applied Logic 156 (2008) 170–182.
[27] Claus P. Schnorr, Zufälligkeit und Wahrscheinlichkeit, Springer, Berlin, 1971.
[28] Glenn Shafer, From Cournot's principle to market efficiency, in: Jean-Philippe Touffut (Ed.), Augustin Cournot: Modelling Economics, Edward Elgar, 2007, pp. 55–95. Can be downloaded from http://probabilityandfinance.com, Working Paper 15.
[29] Glenn Shafer, Vladimir Vovk, Probability and Finance: It's Only a Game!, Wiley, New York, 2001.
[30] Glenn Shafer, Vladimir Vovk, Akimichi Takemura, Lévy's zero-one law in game-theoretic probability, Technical Report arXiv:0905.0254 [math.PR], `arXiv.org` e-Print archive, May 2009.
[31] Jean Ville, Etude critique de la notion de collectif, Gauthier-Villars, Paris, 1939.
[32] Vladimir Vovk, A logic of probability, with application to the foundations of statistics (with discussion), Journal of the Royal Statistical Society B 55 (1993) 317–351.
[33] Vladimir Vovk, Continuous-time trading and the emergence of probability, Technical Report arXiv:0904.4364 [math.PR], `arXiv.org` e-Print archive, April 2009.
[34] Vladimir Vovk, Prequential probabability: game-theoretic = measure-theoretic, in: Proceedings of the 2009 Joint Statistical Meetings, Section on Risk Analysis, Session on Predictive and Prequential Statistics, American Statistical Society, Alexandria, VA, 2009, pp. 3859–3872. Also published as Technical Report arXiv:0905.1673 [math.ST], `arXiv.org` e-Print archive, May 2009.
[35] Vladimir Vovk, Alexander Shen, Prequential randomness, in: Yoav Freund, László Györfi, György Turán, Thomas Zeugmann (Eds.), Proceedings of the Nineteenth International Conference on Algorithmic Learning Theory, in: Lecture Notes in Artificial Intelligence, vol. 5254, Springer, Berlin, 2008, pp. 154–168.
[36] Vladimir Vovk, Vladimir V. V'yugin, On the empirical validity of the Bayesian method, Journal of the Royal Statistical Society B 55 (1993) 253–266.